Word Sense Disambiguation in Text-to-Pictograph Conversion
Gilles Jacobs, Leen Sevens, Vincent Vandeghinste, Inneke Schuurman, Frank Van Eynde

We describe the implementation and evaluation of word sense disambiguation (WSD) in a translation system that converts English text messages into sequences of pictographic images. Text-to-Picto is a text-to-pictograph translation system for Dutch, English, and Spanish used in an online communication platform by people with disabilities who have trouble reading and writing. The translation system relies on the WordNet lexical database in which synsets are populated with pictographs. Many ambiguous words are translated incorrectly by a pictograph which is linked to the wrong word sense due to the lack of WSD in the current system.
The WSD method required for our translation engine must work on general domain text and use WordNet sense inventories. We opted for the gloss-overlap, extended lesk algorithm described by Banerjee and Pedersen (2002). During translation each possible WordNet synset of every content word in the input sentence receives a disambiguation score. This score alongside other parameters are used in a path-finding algorithm to determine the optimal pictograph sequence during translation. This implementation approach is easily generalised to other sense labelling algorithms such as an SVM WSD tool for Dutch (Izquierdo 2015).
In evaluation of the translation output, an improvement over the baseline system without WSD was not obtained. However, we found that WSD works well for ambiguous words for which sufficient pictographs are linked in our lexical-pictorial database.

Banerjee, Satanjeev, and Ted Pedersen. 2002. “An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet.” In Computational Linguistics and Intelligent Text Processing, 136–45. Springer.
Izquirdo, Ruben. 2015. “”DutchSemCor WSD system for Dutch.” RubenIzquierdoBevia.com, October 2015, http://rubenizquierdobevia.com/software/#module7.